

19846  
SEARCH REQUEST FORM

Examiner # (Mandatory): \_\_\_\_\_ Requester's Full Name: \_\_\_\_\_

Art Unit \_\_\_\_\_ Location (Bldg/Room#): \_\_\_\_\_ Phone (circle 305 306 308) \_\_\_\_\_

Serial Number: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Date: \_\_\_\_\_

Keywords (include any known synonyms registry numbers, explanation of initialisms): \_\_\_\_\_

## Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

## STAFF USE ONLY

Searcher: Belcher, C 4994

Searcher Phone #: \_\_\_\_\_

Searcher Location: \_\_\_\_\_

Date Picked Up: \_\_\_\_\_

Date Completed: 09/09-23-99Clerical Prep Time: 5Terminal Time: 20Number of Databases: 1

## Type of Search

\_\_\_\_\_ N.A. Sequence

\_\_\_\_\_ A.A. Sequence

\_\_\_\_\_ Structure (#)

\_\_\_\_\_ Bibliographic

\_\_\_\_\_ Litigation 1

\_\_\_\_\_ Fulltext

\_\_\_\_\_ Procurement

\_\_\_\_\_ Other

## Vendors (include cost where applicable)

\_\_\_\_\_ STN

\_\_\_\_\_ Questel/Orbit

\_\_\_\_\_ Lexis/Nexis

\_\_\_\_\_ WWW/Internet

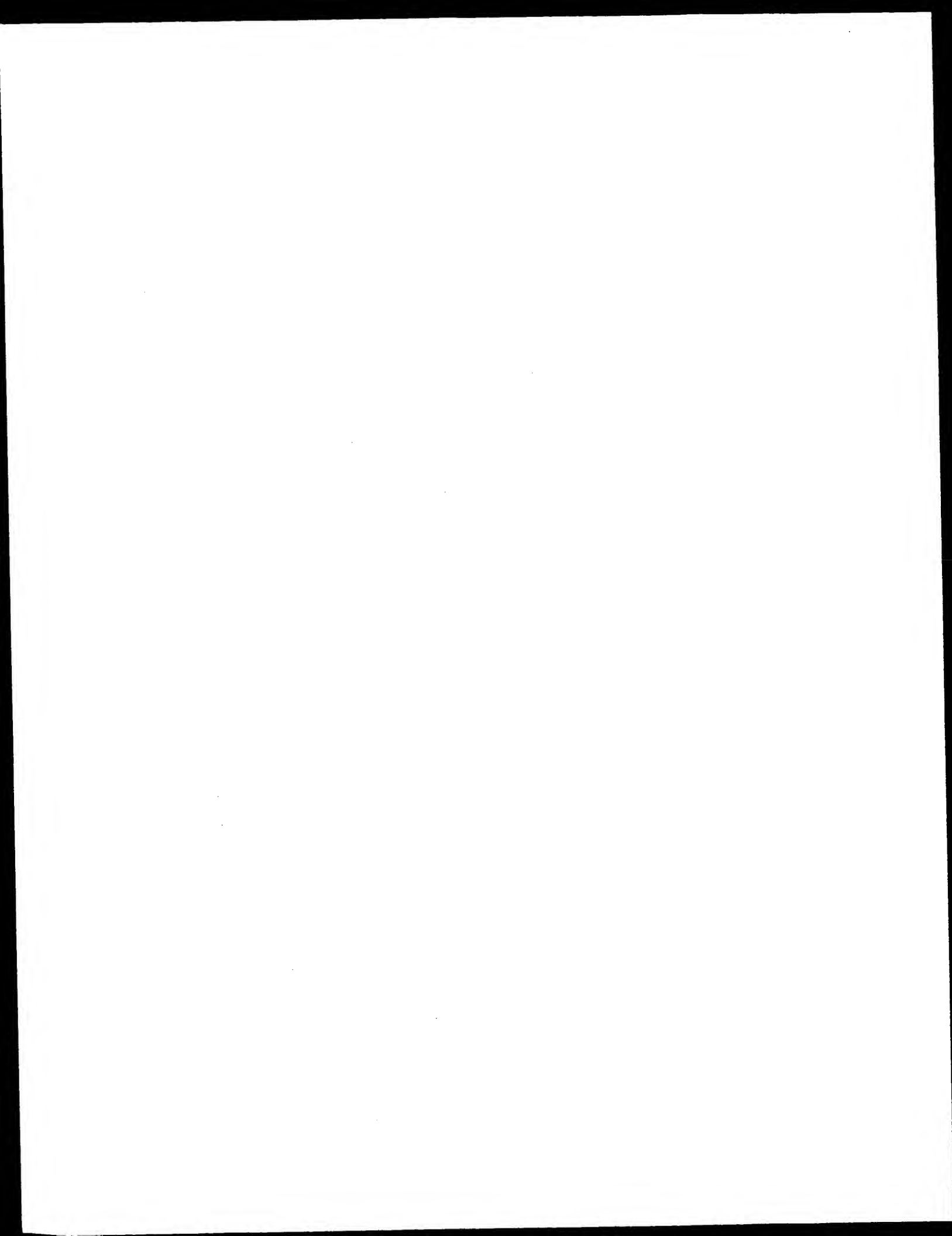
\_\_\_\_\_ In-house sequence systems (list)

\_\_\_\_\_ Dialog

\_\_\_\_\_ Dr. Link

\_\_\_\_\_ Westlaw

\_\_\_\_\_ Other (specify) Compu Gen



Thu Sep 23 10:37:29 1999

us-08-671-757-1.rge

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd

Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 1999, 16:53:41 ; Search time 99.9 Seconds

(without alignments)  
827.708 Million cell updates/sec

Title:	US-08-671-757-1
Perfect score:	76

Sequence: 1 ATGCTCGAGGTCGAAAGCAAGATG 26

Scoring table: IDENTITY\_NUC

Scoring table: IDENTITY\_NUC

679419 seqs, 1590154680 residues

Database

GenEmb1: \*

1:	gb_ba1: *
2:	gb_ba2: *
3:	gb_om: *
4:	gb_ov: *
5:	gb_pat: *
6:	gb_ph: *
7:	gb_pl1: *
8:	gb_pl2: *
9:	gb_pt1: *
10:	gb_pt2: *
11:	gb_pt3: *
12:	gb_ro: *
13:	gb_st: *
14:	gb_sts: *
15:	gb_sy: *
16:	gb_un: *
17:	gb_v1: *
18:	em_fm: *
19:	em_hb: *
20:	em_hun1: *
21:	em_hun2: *
22:	em_in: *
23:	em_om: *
24:	em_ov: *
25:	em_ov: *
26:	em_ph: *
27:	em_ph: *
28:	em_pl: *
29:	em_to: *
30:	em_sts: *
31:	em_sy: *
32:	em_un: *
33:	em_v1: *
34:	gb_hg1: *
35:	gb_hg2: *
36:	gb_in1: *
37:	gb_in2: *
38:	em_ba2: *
39:	em_ba2: *
40:	em_hun3: *
41:	em_hun4: *
42:	gb_pt4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query Match	Length	DB	ID	Description
No.						
1	26	100.0	26	5	A59060	A59060 Sequence 1

C	2	18.6	71.5	2657	2	AE0826661
C	3	18.2	70.0	15585	2	AE000078
C	4	18.2	70.0	131993	17	OPU759330
C	5					
C	6	18	69.2	91539	8	ATAC00428
C	7	18	69.2	100000	8	AP000118
C	8	18	69.2	100000	9	AP000165
C	9	18	69.2	68109	9	AP000315
C	10	18	69.2	24518	9	AP000316
C	11	18	69.2	69013	9	HS138A5
C	12	18	69.2	1120	12	AF0098989
C	13	18	69.2	1312	12	RNO132222
C	14	18	69.2	1162	17	RNU61078
C	15	18	69.2	1657	17	PNVEP
C	16	17.6	67.7	30	1	HS796E4
C	17	17.6	67.7	5092	1	ABO11599
C	18	17.6	67.7	4400	1	HPCYTXXO
C	19	17.6	67.7	623	1	HPU07145
C	20	17.6	67.7	628	1	HPVACA26
C	21	17.6	67.7	628	1	HPVACA3
C	22	17.6	67.7	67.3	1	HPVACA31
C	23	17.6	67.7	1133	1	HPY66V67
C	24	17.6	67.7	4200	1	S72444
C	25	17.6	67.7	732	2	AF0334402
C	26	17.6	67.7	706	2	AF0336511
C	27	17.6	67.7	4196	2	AF0496622
C	28	17.6	67.7	4286	2	AF0496627
C	29	17.6	67.7	4207	2	AF0496632
C	30	17.6	67.7	4162	2	AF0496631
C	31	17.6	67.7	4165	2	AF0496638
C	32	17.6	67.7	4216	2	AF0496646
C	33	17.6	67.7	4751	2	AF0496648
C	34	17.6	67.7	4239	2	AF0496650
C	35	17.6	67.7	4233	2	AF0496652
C	36	17.6	67.7	721	2	AF0496652
C	37	17.6	67.7	89603	2	AF091815
C	38	17.6	67.7	56631	8	AC007659
C	39	17.6	67.7	131389	11	HS702M17
C	40	17.6	67.7	15480	17	BYVUA
C	41	17.6	67.7	6746	17	CLAB3PBH
C	42	17.6	67.7	110000	34	HS437G10
C	43	17.6	67.7	110000	34	HS437G10
C	44	17.6	67.7	172008	35	AC007185
C	45	17.6	67.7	193301	35	AC0073932

## ALIGNMENTS

```

RESULT      1
LOCUS       A59060
DEFINITION  A59060      26 bp      DNA
ACCESSION   A59060      Sequence 1 from Patent EP0752473.
NID         g3714495
VERSION     A59060.1    GI:3714495
KEYWORDS
SOURCE      .
            unidentified.
ORGANISM    unidentified
            unclassified.
REFERENCE   1 (bases 1 to 26)
AUTHORS     Suerbaum,S. and Tadjine,A.
TITLE       Cloning and characterization of H. pylori fliA gene: production of
            flagella-free strains
JOURNAL     Patent: EP 0752473-A 1 08-JAN-1997;
            PASTEUR INSTITUTE (FR)
COMMENT     Other publication CA 2180473 970105
            Other publication AU 583596 970116
            Other publication FR 276360 970110.
FEATURES
SOURCE      location/Qualifiers
            1..26
            /organism="unidentified"
            /db_xref="taxon:32644"

```

```

BASE COUNT      9 a      5 c      8 g      4 t
ORIGIN
Query Match      100.0%; Score 26; DB 5; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 ATGCCTCGAGTCGAAAAGCAAGATG 26
        |||||
Db      1 ATGCCTCGAGTCGAAAAGCAAGATG 26

RESULT 2
AF082661/c      2657 bp      DNA      BCT      11-NOV-1998
LOCUS      Meliothermus ruber proba operon, complete sequence.
ACCESSION      AF082661
VERSION      AF082661.1 GI:3859497
KEYWORDS
SOURCE      Meliothermus ruber.
ORGANISM      Bacteria; Thermus/Deinococcus group; Thermus group; Meliothermus.
REFERENCE      1 (bases 1 to 2657)
AUTHORS      Yaklichkin,S.Y., Zimina,M.S., Yurchenko,Y.V., Hromov,I.S. and
              Neumiyakin,L.V.
TITLE      Molecular cloning and sequence analysis of the proA gene from
              thermophilic eubacterium Thermus ruber
              Unpublished
              2 (bases 1 to 2657)
              Yaklichkin,S.Y., Zimina,M.S. and Neumiyakin,L.V.
              Direct Submission
              Submitted (06-AUG-1998) Center of Biotechnology and Molecular
              Diagnostics, Institute of Molecular Genetics, RASc, Kurchatov
              Square, Moscow 123184, Russia
              3 (bases 1 to 2657)
              Yaklichkin,S.Y., Zimina,M.S. and Neumiyakin,L.V.
              Direct Submission
              Submitted (10-NOV-1998) Center of Biotechnology and Molecular
              Diagnostics, Institute of Molecular Genetics, RASc, Kurchatov
              Square, Moscow 123184, Russia
              Sequence update by submitter
              On Nov 11, 1998 this sequence version replaced gi:3514027.
              Location/Qualifiers
              1..2657
               /organism="Meliothermus ruber"
               /strain="40"
               /db_xref="taxon:277"
               1..2657
               /gene="proA operon"
               61..1233
               /gene="proB"
               61..1233
               /gene="proB"
               /codon_start=1
               /transl_table=11
               /product="gamma glutamyl kinase"
               /protein_id="AAC72812.1"
               /db_xref="PID:g3859498"
               /db_xref="GI:3859498"
               /translation="MRNAPRSVTTTQAMHGRLPTAOTHRRLVYKVSATLSGPGRO
               HOLAIQAQVAALRAEGRVIVSSGAQGTGQKLTGTEKFRSPGKALAAVOPITLM
               LIMEQAFSWYDLKVAQVLLTIEDLAHRRYINARQLETLEWGIPIINENDPVVE
               EIKFGDNQSLASLIVGADLLILSLRIEALYEDAPRTPEAQPIYVERVAGVLR
               MAGSPNVGTGSKSLIAEKAQAAGIPIILLPGTRPOSIAALQGAAGVGLTFAG
               QRRYSGRKIMLYOLPKPGEEVVDAGAKAIRGGASTLPAGILLEVGGQVGEAVVC
               LDEQNLIGVGLVNTSAELAKIRKRTREIALGKYNDEAIHRDYPALAELE"
               1239..1244
               /gene="proA operon"
               /note="putative"
               1249..2502
               /gene="proA"

```

```

CDS
1249..2502
/gene="proA"
/EC_number="1.2.1.41"
/codon_start=1
/transl_table=11
/product="gamma-glutamyl phosphate reductase"
/protein_id="AAC72811.1"
/db_xref="GI:3514028"
/db_xref="PID:g3514028"
/translation="MVTSPDLKAYQAARAAARALSTASPRAKNTALLIAKLEAQ
EALPANREDLEAAQAGLSKARLDRLRDLRTGLQGVAMPDPVGEIQLQ
IRPNGLGVGRVPLGVIGFYSPRYATVEASALCLKANALLRGKRAMSNAL
VGLMGAALQERAGLPREAIOLVPTTDSALILEMCHLADLDLILPRGREGILEVQEA
RIPLYATREGINHLFVDESDTPGCAVQIALNGTORPSCNSLEKYLHQRARAPVL
MLAQMOKAGVELEGDEATCALIPARPAIPEMDOTELDILITKAYNSIEEALHIA
RYGSHHTAEICTNHNHARFRLREVDSALVLNAPSFRFNGFQGLGAEIGISTSKLH
AYGPMGVKEITTTTFVALGSGOYRD"

BASE COUNT      490 a      899 c      866 g      402 t
ORIGIN
Query Match      71.5%; Score 18.6; DB 2; Length 2657;
Best Local Similarity 84.0%; Pred. No. 99;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 ATGCCTCGAGTCGAAAAGCAAGAT 25
        |
Db      634 AAGCCTCGATGTCGAGAGCAAGAT 610

RESULT 3
AE000078/c      15586 bp      DNA      BCT      12-DEC-1997
LOCUS      Rhizobium sp. NGR234 plasmid pNGR234a, section 15 of 46 of the
DEFINITION      complete plasmid sequence.
ACCESSION      AE000078 U00090
NID      92182439
VERSION      AE000078.1 GI:2182439
KEYWORDS
SOURCE      Rhizobium sp. NGR234.
ORGANISM      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
              Rhizobiaceae; Rhizobium.
              1 (bases 1 to 15586)
              Freiberg,C., Fellay,R., Bairoch,A., Broughton,W.J., Rosenthal,A.
              and Perret,X.
              Molecular basis of symbiosis between Rhizobium and legumes
              Nature 387 (6631), 394-401 (1997)
              2 (bases 1 to 15586)
              Freiberg,C.
              Direct Submission
              Submitted (22-NOV-1996) Genome Analysis, Institute for Molecular
              Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
              3 (bases 1 to 15586)
              Freiberg,C.
              Direct Submission
              Submitted (12-DEC-1997) Genome Analysis, Institute for Molecular
              Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
              Update by submitter
              Location/Qualifiers
              1..15586
               /organism="Rhizobium sp. NGR234"
               /plasmid="pNGR234a"
               /strain="NGR234"
               /db_xref="taxon:394"
               <1..>15586
               /organism="Rhizobium sp. NGR234"
               /db_xref="taxon:394"
               /clone="pXB182"
               complement(82..130)
               /note="nod box-like sequence"
               /note="region: nod box 10"

```

```

misc-feature

```

```

gene
/feature=not_experimental
complement(178..870)
/gene="y41c"
complement(178..870)
/gene="y41c"
/feature="hypothetical 25.8 kd protein"
/codon_start=1
/transl_table=11
/product="y41c"
/protein_id="AAB91698.1"
/db_xref="PID:92182440"
/db_xref="GI:2182440"
/translation="MYREBAFNAPARASGYDLDAAAYGOOSNGRGEPRFGDIER
LICTGTGKVLPGTPTVTETAMRNKIPKKEIDTILIELNVLVRRRVMAGIRSK
PASEIMGVYRALDGPLAPHCASRAFCDDCNDPVTQSVLYCHRRKATIGYGA
AMLRNASRVELSDVSRRIIAOHQLRISQRPRAOYCPASRRNRHNETEMGGLRFGII
VRIDRHTOSLYN"
complement(812..2800)
/gene="y41d"
complement(812..2800)
/gene="y41d"
/feature="probable monooxygenase"
/codon_start=1
/transl_table=11
/product="y41d"
/protein_id="AAB91699.1"
/db_xref="PID:92182441"
/db_xref="GI:2182441"
/translation="MAHSSDLALFRNRHIAELPESANVPTLLLLYOPTGKEYMLTPPE
LPKSVGNDNGSLAELOTEIRNALDAILICHSAPYKODLSEELIRMLTSE
AEPIDPEADMDHKLRRYAGAIIDPVSLPKEFVLLIGAMSGVAAIRLGLTISY
IOVKPDSGTGVMHAHHYPGCGVDTPGHIXSYTPASGNMSTFPPLQKEIDYNNRVA
DGETESTRTTECLVTRIDESTLTHSRRLRPLNGTETLVTNVLSAVGFTTPKP
NLSGIJRNEDGVVHTSKMDPEVALDGRKRAVINGASAMQVPAIADRGALTIPOKS
ROMVAPPEKQKVPPEPMQELFREVPHEMLRLSLMIDSEVHEALQDPAMPHPD
RSVANAVDRDEATYOTIEGOLAPDLIAVIPSYPFGKMLDNGWRTLLKPIY
TLVDGAARVGRSIIHAIDDETHADLIIVASGYDTTIRLLPVEVIGRNQRTYRDWD
DDDCATIGTYVAFNPFEMLYGPTALGHRGSITITIEQIDVLSVLQOMKEKRY
AEVECRDITYCHNRKIOOHQOMTWSHEGMSSTFERNDRGRITVNSPWLVDYNLKE
ADDDYRTMPQVDLSLETSGVREGVORPGSRRLRRPS"
3544..3951
/gene="y41e"
3544..3951
/gene="y41e"
/feature="hypothetical 15.4 kd protein (possibly
fragmentous); similar to y41L,y42a"
/codon_start=1
/transl_table=11
/product="y41e"
/protein_id="AAB91700.1"
/db_xref="PID:92182442"
/db_xref="GI:2182442"
/translation="MOERLALADTTPPAMRSDGVAOPIDFLIASLRTAMKDGATRPD
RPIYAKGRRRPPVLYKATADLRSMFPAEPWRGSEFLSLAQEYEGDYPDKLLRTL
GRLLKVMSPQADALICTLNLKELPQOIAIRPH"
3544..3951
/gene="y41f"
/feature="insertion sequence homolog (ISH) 8g: region
containing putative IS element-derived ORF (y41e)"
/feature="Region: ISH8g"
complement(4460..5377)
/gene="y41f"
complement(4460..5377)
/gene="y41f"
complement(4460..5377)
/gene="y41f"
/feature="hypothetical 34.1 kd protein"
/codon_start=1
/transl_table=11
/product="y41f"
/protein_id="AAB91701.1"
/db_xref="PID:92182443"
/db_xref="GI:2182443"
/translation="MSGALLMAKKAQORISSALVYGTNNVDAELKLAOKLHGVRDY
VQMHQDRCKRGISVLEPGYVAFAPQKTLPLVIGLGETSKCTLVDRSRETLGDI

```

---

```

misc-feature
/feature="insertion sequence homolog (ISH) 11c: region
containing putative IS element-derived ORF (y41g)"
/feature="Region: ISH11c"
complement(5839..6111)
/gene="y41g"
complement(5839..6111)
/gene="y41g"
/feature="hypothetical 10.5 kd protein (possibly
fragmentous); similar to y44f and y44p"
/codon_start=1
/transl_table=11
/product="y41g"
/protein_id="AAB91702.1"
/db_xref="PID:92182444"
/db_xref="GI:2182444"
/translation="MLDLYREHYLDFPTLAEKILRHRLAVSKETLRQIMAGIIN
ALRRERKRVRPQGRGRCFGLLIQIDSGPDTKKSIVLPILSTAR"
6294..7013
/gene="y41h"
6294..7013
/gene="y41h"
/feature="hypothetical 26 kd protein precursor"
/codon_start=1
/transl_table=11
/product="y41h"
/protein_id="AAB91703.1"
/db_xref="PID:92182445"
/db_xref="GI:2182445"
/translation="KRTVAMLLAAVVAASSTLAVNFCQSKNRNCTGSGFLVD
IPGDSHKLKIDFCGYVDLNGWQNTKEITDGSIPPLQDPFGSPWEDGYIRAY
IHDWCDCRIVRTKETHREVDYTMASGLEPKAKLLFYAVYAGPBMGYLVPEKCA
ACKNCIOMTGKRAAFVOLRGELADQSSAGELKATIKATIDLKERSGDATLDELMAID
EAHPKQTLRDQRPAGDITIK"
complement(7186..9297)
/gene="y41i"
complement(7186..9297)
/gene="y41i"
/feature="hypothetical 76.2 kd integral membrane protein"
/codon_start=1
/transl_table=11
/product="y41i"
/protein_id="AAB91704.1"
/db_xref="PID:92182446"
/db_xref="GI:2182446"
/translation="HRRSTEGTWTVDGSKPTFEIGIMSGAISGATISAGVDFELI
ALDAMWKARAEAGAPLPEYDVKLALSGASAGATTAIGVIAAGREAPVFPSPAG
SONIRFTLRLRSWVTSPTLIVSPGSDLSLELDAGRPVTSVLNANVLTAGAEB
LEATGTLSPRAVASSLHLYMMLSLRVPYAIHNGGOYMMTADHVRHVVGIGT
WKPTSPFPADTDSGTISATSLFGAAGASTREPEULAAANALAGAPIGISPVIA
TATSOYAKKAPFISDSQSLRPIPTWPAHMAHSAODPSPFSVSDGLINNDEPV
RTTLAKDPPRPNERNKADRATVITATPPEGPPPLDGGEPPLVSLTARRVYARLQ
QVRFKPDQLAAAGTSHREMIISPRVPSTPGSEEREETISLISGLLGGGGVLE
AFRDHDYQGRNRCQFLMRHLITIDKNQTLMPGGAERNAVTSKTLSDGSHDYY
PIIPLVDGLDPEVPYRVARIDENAFALVYKIEARLVAARAVSTETTSARKMLGL
NFLLAGRARIIDYRLTLLQELVRDQIEGPIPAALDRPDEVAVIAALLDPAFLD
RKEAGIARTTKIDTISLLEITSLAGAGANQVLAWTRIDERSLITIVSRRSFL
ATLLQGRSPARLFAKPVYDRK"
7567..15386
/organism="Rhizobium sp. NGR234"
/db_xref="taxon:394"
/clone="BX81661"
complement(9346..11136)
/gene="y41j"
complement(9346..11136)
/gene="y41j"
/feature="hypothetical 65.5 kd protein"
/codon_start=1
/transl_table=11

```

```

/product="y41j"
/protelin_id="AAB91705.1"
/db_xref="PID:92182447"
/db_xref="GI:2182447"
/translocation="MKRKSFWPILLGFTVLVAAGLYVVMESVDLPDPYKRVKM
LEONMSOSORGMHMHADOGTVFESMPEWMLAEOTPTLTAGPELSDDYDRGFI
TADSSGLPVGFAGHDLDVDPKTAQWPVNATGPKPLTVGLTACHTGRTYKKAAM
VDCGPAITDLGKFRKASGLAFPTRPAPRPFDRFANAVIGPQADKARAVLKQDKY
LAGRIEVDLDKVAEKSTIEGGRIDLNRIGNOVFSIDLERPEYVAQSAFVAFPH
IWDTSFDMVQINASIMQPMVRNAGALVRAFINLTKSEQPLFASVAKVDITFIEQ
QLAGKOPTAENGTEGLRPMPNSNLFSGIDTKLATGAAVYADRCOGCHLPPVGSFG
MECKHWTNENSAGERLYRVPILINVENIGTDPALQASOMERKYLPESELGIDDSGSA
LGALVAKTAARWYDNOTPPVPAEQREINNGNONGIQAPLAYKARPLDGIWATPPEHL

Query Match      70.0%; Score 18.2; DB 2; Length 1586;
Best Local Similarity 87.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2  TGCCTCGAGTCCGAAGAAGA 24
| | | | | | | | | | | | | | | |
Db      2323 TACATCCAGGTCCGAAGAAGA 2301

RESULT  4
LOCUS   OP075930      131993 bp      DNA      circular      VRL      06-MAR-1998
DEFINITION  Orygia pseudotsugata nuclear polyhedrosis virus complete genome.
ACCESSION  U75930
MID      92934903
VERSION   075930.1  GI:2934903
KEYWORDS
SOURCE    Orygia pseudotsugata nuclear polyhedrosis virus.
           Orygia pseudotsugata nuclear polyhedrosis virus.
           Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
           Nucleopolyhedrovirus.
REFERENCE  1 (bases 1 to 131993)
           Ahrens,C.H., Russell,R.L., Funk,C.J., Evans,J.T., Harwood,S.H. and
           Rohmann,G.F.
           The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
           polyhedrosis virus genome
           Virology 229 (2), 381-399 (1997)
           2 (bases 1 to 131993)
           Rohmann,G.F.
           Direct Submission
           Submitted (23-OCT-1996) Oregon State University, Agricultural
           Chemistry, Corvallis, OR 97331-7301, USA
           3 (bases 1 to 131993)
           Rohmann,G.F.
           Direct Submission
           Submitted (06-MAR-1998) Oregon State University, Agricultural
           Chemistry, Corvallis, OR 97331-7301, USA
           Sequence update by submitter
           On Mar 6, 1998 this sequence version replaced GI:1911246.
           similar to Autographa californica nuclear polyhedrosis virus
           (AcMNPV) complete genome: Genbank Accession Number L22858.
FEATURES
     source
     CDS
     1..131993
     /organism="Orygia pseudotsugata nuclear polyhedrosis
     virus"
     /db_xref="taxon:10450"
     complement(123..947)
     /note="ORF1: PK-1, AcMNPV ORF10 homolog"
     /codon_start=1
     /product="protein kinase 1"
     /protein_id="AAC59000.1"
     /db_xref="PID:91911247"
     /db_xref="GI:1911247"
     /translation="MDALQSLSRVADCAVIAPVNVGRFGKMDVHHRPTSKLFL
     RYIAHSFSADINVDLMSDHPFVDMYFCYSPTMAIVMDYVPCPDLETLQIG
     GALDNALVNIYRQLCDALNDLHNTATGYIHNQVLENLVYGARVYVLCGYGCKRE
     HSPVHGTEYFSPKIRRHNYARVDWYAVGLAYKLLTGSRHPEERSADVEIDLAS
     MRROQYNDPALKNVRNLMADVFECULRFNFECSRIDYKQIAKHSFLASRHYI"

     CDS
     946..2370
     /note="ORF2: AcMNPV ORF9 homolog"
     /codon_start=1
     /product="1629-capsid"
     /protein_id="AAC59001.1"
     /db_xref="PID:92934904"
     /db_xref="GI:2934904"
     /translation="MMEQIQSVRSYLINQHNALNAGFFLORVAGPEASVGRNVD
     RAURLNQAVLDLKLAEIDYADTAQMDQEPASSRHRPATLNRMLLIGVQDDPAR
     RINLSVATLEALRLVDVNDADAVNLSGDFEYESKYSISYQTFQPTAGASQQTQ
     TSLRPQTSLSPQTOASLDPOOTQEPENVSPPSVHTTAPILPOTTOPPATDTSRP
     SDEVVYVQGERAVPTDRFPKPPKPEHLKSRSSVATNAGATPVPAPPPPSGAD
     VTSMPPPPPPSADVTSMPPPPPMVDLASMPPPPPPPMVDLASMPPPPPINNA
     INLLIDAWAETNKNAGDNRSLDDQIQOGTLTKTQPDAGPATDPSLSTLRQ
     GKTLKURKLEDSSTQTLTKVDYDTDKTKTLKNTVINDRISQOEQEKRLDTIT
     KRPAVEHRTDGNSTGNSSDDMRD"
     complement(2533..3270)
     /note="ORF3: ph. AcMNPV ORF8 homolog"
     /codon_start=1
     /product="polyhedrin"
     /protein_id="AAC59002.1"
     /db_xref="PID:91911249"
     /db_xref="GI:1911249"
     /translation="MPDYASYRPTIGRTYYVDNKKYKNIKNGSVTKNAKKHLEHDE
     KHDPLDHYVVAEDPFLGSGKOKLTFKEIRNVKPDYTKLILNNSGKFLLETTRF
     VDSFPYVDQEDVDFVLYNNKPTLRNCTYFLAQHAIKMDQDYPHEVIVIEVSY
     EILIEVSLVKEVKEFADPAFLFTGPAY"
     complement(3350..3541)
     /note="ORF4"
     /codon_start=1
     /product="unknown"
     /protein_id="AAC59003.1"
     /db_xref="PID:91911250"
     /db_xref="GI:1911250"
     /translation="MRKNAGASQYPEHLTVKRRHRLQSLCWCDDQSRWDSHSAKGL
     MEFHMFCCIRIYQKKIKH"
     complement(3698..4093)
     /note="ORF5"
     /codon_start=1
     /product="unknown"
     /protein_id="AAC59004.1"
     /db_xref="PID:91911251"
     /db_xref="GI:1911251"
     /translation="MGASQNSEOKILIKLFYRMSQTAGALDDEKDLHCLYDIERFVGA
     HLNRKADDAKARKKCAEKALKRVELIADRRHMLEAAPACADGDRWSTLSRAQLDDI
     ARKEIVDIRIHRQLKODSLKODRLKR"
     complement(4125..4739)
     /note="ORF6: 1ef-2, AcMNPV ORF6 homolog"
     /codon_start=1
     /product="late expression factor 2"
     /protein_id="AAC59005.1"
     /db_xref="PID:91911252"
     /db_xref="GI:1911252"
     /translation="MERVWNPAAAGIDGLKRSFTYLVDPHDYGVLTSLPYFEGCLE
     VRMSGRLLALAAKRPQEPQAVRRFPQRSRNVCCLACADGAOSLKLVAARVSMF
     PCKSKTYMADLSSAPRGNMYRKRFENCYLANVITCTKCTACCLIGALLHYRMAKCV
     GEVTHLLIKAQDYVYKPCNCAKMKVTKLCPQASMKGILNCFN"
     complement(4742..4975)
     /note="ORF7: AcMNPV ORF5 homolog"
     /codon_start=1
     /product="unknown"
     /protein_id="AAC59006.1"
     /db_xref="PID:91911253"
     /db_xref="GI:1911253"
     /translation="NNRPMTNMTAAVTVDYDEQRLRELNSLRSVHELCTRSAFGPD
     CNRELEAGDRAPAVIYKAAANGQSSSLICKV"
     5020..5460
     /note="ORF8: AcMNPV ORF4 homolog"
     /codon_start=1
     /product="unknown"
     /protein_id="AAC59007.1"
     /db_xref="PID:91911254"
     /db_xref="GI:1911254"

```

CDS

/db\_xref="GI:1911254"  
/translation="MPGARFVRSRLTQFEKENVAHVHILMGLRALIDGKYSADV  
RREFSLRNITLVSNCAVMNVNOYCAPDPAIDMHOPITIFRYVCNCHAMADVPAPDDHS  
IARYLLAECAGVELTDHPDVEDGETEGVENLLEVRINACGD."  
complement(5669..6151)  
/note="ORF9; ptp-2"  
/codon\_start=1  
/product="protein tyrosine phosphatase 2"  
/protein\_id="AAC59008.1"  
/db\_xref="PID:g1911255"  
/db\_xref="GI:1911255"  
/translation="MEASQIDHEVFVGGITGDNEMMLFIENIRIASYSTIDSQA  
PIRMALGIPVDSDIHVCEAPATCALPNMALPDYMRIGBSKRYLVHCTAGASR  
SAAALVAYLMRSOMPYPQDALNAVNOSKRRYAINDHVFLARTRCSYRFVNDLKPOLY  
"  
complement(6129..6791)  
/note="ORF10; ptp-1, AcMNPV ORF1 homolog"  
/codon\_start=1  
/product="protein tyrosine phosphatase 1"  
/protein\_id="AAC59009.1"  
/db\_xref="PID:g1911256"  
/db\_xref="GI:1911256"  
/translation="MTLSRRRAVVIYRTSRNCGAGCTNGVRRROYHLPSVARAM  
PDRNHETACGAVIECTRLLCFEVPNALPEPVYISDEPRHTASTASVARSAGAVI  
DLTNIRAYDAQWKMGLTYKRIYGRVPPDDIVAFITVDPFRRCPTMLGVV  
HHTIGNSGYGLCYRYMYERLGVSPTDIARFETANGHKITEKITNLQDLLARKHYRG  
PN"  
6819..7814  
/note="ORF11; AcMNPV ORF11 homolog"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAC59010.1"  
/db\_xref="PID:g1911257"  
/db\_xref="GI:1911257"  
/translation="MSLSKLLVYAAYGSYNLPEDHRGESYHLRYIVEHRLNTVSN  
ASGVRDITACLINSGLCPDVARNLDIDSVEARLSASFRCGDATGLCADMOHALA  
DIDNRAPLARVRGRANFALDAIDIPDTNNLOGITGERFMHPRCGLARVALA  
DPDIRADGWKHRCFVLTYMHLYACGANPASATRPADYAKHIGNEGNCAPIIAA  
LVRFAALVIGHGTIBROKNALEFDLLCCVNASDIDADCYDCVNFKYATOXKKTKM"  
complement(8180..9142)  
/note="ORF12; AcMNPV ORF13 homolog"  
/codon\_start=1  
/product="unknown"  
/protein\_id="AAC59011.1"  
/db\_xref="PID:g1911258"  
/db\_xref="GI:1911258"  
/translation="MFSSMGWMNADAEOVNAEFDEOAYRRIYADORAHSIDLRYDF  
RCHPEYFKFYVLTDETAGSCSYVDFCKGLKISHDLLQRCPFNQHOVROLNTELVGAP  
PADPDISGIS.FATRHGLVOLLOOFSEFANKNYLAIYANDKHDPNDILIKTEIVLVN  
KTLTNLNDKRTISAHKFSFKLEVARGPEOFORLOTLDILNKALOCAATPYAVVPR  
DYTLNHPILAIFSGRVDRGVTOIARAQOEHFREKRIKEFPEGNDVRSRAPNPILA  
VHCICEEFANGHGKIRLRPKRYIEVDCANVAKD.VKKAILINKI"  
complement(9094..9825)  
/note="ORF13; left-1, AcMNPV ORF14 homolog"  
/codon\_start=1  
/product="late expression factor 1"  
/protein\_id="AAC59012.1"  
/db\_xref="PID:g1911259"  
/db\_xref="GI:1911259"  
/translation="MAPCKYTPERVOMMDAIYVDSRRLAFMTDRPRRWYHANFPDS  
AGAEFAATVINESIDVHVKLDEGGREWITADFCDACKDELKLVNVAATAFMLPF  
EKGEDAVOIRNESGNRGFHLMKFCKGFMDAPKLF.RHFWNVNKPQAPKLVSDIRG  
SFACVCAWRVYITGAARDLVTRYMPDVDVPCFNANKQIRAFESYNYKGDYSRCL  
TQQIQARKACSAGCLAGTTPITSK"  
9893..11362  
/note="ORF14; egt, AcMNPV ORF15 homolog"  
/codon\_start=1  
/product="ecdysteroid UDP-glucosyltransferase"  
/protein\_id="AAC59013.1"  
/db\_xref="PID:g1911260"  
/db\_xref="GI:1911260"  
/translation="MVFLIIATLIATLTAAGAASILIADVLPTRYKSHVVYRAYVHALVK

CDS

/db\_xref="GI:1911254"

CDS

/db\_xref="GI:1911254"

CDS

/db\_xref="GI:1911254"

Query Match	70.0%	Score 18.2	DB 17	Length 131933
Best Local Similarity	87.0%	Pred. No. 1.2e+02		
Matches	20	Conservative	0	Mismatches 3
				Indels 0
				Gaps 0
QY	3	GCCTCGAGGTCGAAGAAGCAAGAT	25	
Db	31011	GCCTCGGCGCTCAAAAAGCAAGAT	31033	

	RESULT	5
ATAC004261		
LOCUS	DNA	PLN 02-FEB-1999
DEFINITION	Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence,	
	complete sequence.	
ACCESSION	AC004261	
VERSION	g3402695	
KEYWORDS	AC004261.1	GI:3402695
SOURCE	Atg.	
ORGANISM	thale cress, thaliana	
	Arabidopsis thaliana	
	Brassicaceae: Brassicaceae: Eumycorrhizal: Tracheophyta:	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 91539)  
Rounsley, S.D., Kaul, S., Lin, X., Ketchum, K.A., Crosby, M.L.,  
Braddon, R.C., Sykes, S.M., Mason, T.M., Kerlavage, A.R., Adams, M.D.,  
Somerville, C.R. and Venter, J.C.  
Arabidopsis thaliana chromosome II BAC T3k9 genomic sequence  
Unpublished  
2 (bases 1 to 91539)  
Rounsley, S.D. and Lin, X.  
Direct Submission  
Submitted (05-MAR-1998) The Institute for Genomic Research, 9712  
Rockville, MD 20850, USA. rounsley@tigr.org

REFERENCE	3 (bases 1 to 91539)
AUTHORS	Rounsley, S.D.
TIME	Direct Submission
JOURNAL	Submitted (07-AUG-1998) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
REFERENCE	4 (bases 1 to 91539)
AUTHORS	Lip, X
TIME	Direct Submission
JOURNAL	Submitted (02-FEB-1999) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT	On Aug 7, 1998 this sequence version replaced g1:333494.

The Institute for Genomic Research  
9712 Medical Center Dr.  
Rockville, MD 20850, USA  
e-mail: xlin@igmr.org  
BAC clone T349 is from Arabidopsis chromosome II and is contained  
in the YAC clone C11C108.  
The orientation of the sequence is from SP6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of three methods: Gene  
prediction programs including GNAI (available by anonymous ftp  
from [athur.epm.ornl.gov/](http://athur.epm.ornl.gov/)), GeneIndex (Phil Green, University of  
Washington), GenScan (Chris Bure,  
<http://genomic.stanford.edu/~chris/GENSCAN.html>), and NetPlantGene  
(<http://www.cbs.dtu.dk/nebgenie/cbnebgene.html>), searches of the  
complete sequence against a peptide database and the Arabidopsis  
EST database at TIGR (<http://www.tigr.org/tdb/ar/ar.html>).  
Genes with similarity to other proteins are named  
for their annotation. Genes with similarity to other proteins are named

FEATURES  
source

[illegible]



RESULT	6
AP000118	

DEFINITION	PRI
Homo sapiens genomic DNA of 21q22.1, GART and AML related, SLC5A3-14A4 region, segment 1/8, complete sequence.	21-MAY-1999
ACCESSION	
AP000118	
IID	
54720007	

SEQUENCE	AP000018.1	GI:4730887
KEYWORDS	HTG.	
SOURCE	Homo sapiens DNA.	
ORGANISM	Homo sapiens	

**REFERENCE**  
**AUTHORS**  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (sites), ...  
Hatori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.  
Published Only in Database (1997) In press  
2 (bases 1 to 10000)  
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.  
Direct Submission  
Submitted (15-APR-1999)

**COMMENT**

sequence is submitted by Human Genome Sequencing in ARS project of JST.

Japan Science and Technology Corporation (JST)  
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.ais.tokyo.jst.go.jp/HGS/>) or send email to [webmaster@www.ais.tokyo.jst.go.jp](mailto:webmaster@www.ais.tokyo.jst.go.jp).

location/Qualifiers

1

STS

STS

STS

STS

STS

```

2222.3..2223.9
/note="WI-13874:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4574994"
82240..82341
/note="SHGC-52061:The location is between each flanking

```

```
site of PCR primers."
/db_xref="GDB:6464503"
82240. .82489
```

BASE COUNT	ORIGIN
26820 a	19908 c
	21667 g
	31605 t

query match	69.28;	Score 18;	DB 9;	Length 100000;
Best Local Similarity	80.88;	Pred. No. 1.5e+02;		
Matches	21;	Conservative	0;	Mismatches 5;
				Indels 0;

QY	1	ATGCTCGAGCTCGAAAAGCAAGATG	26
Db	85087	ATGCCCCCAGGGTGTTAAGCAAGATG	85112

RESULT	7
LOCUS	AP000165
DEFINITION	AP000165 100000 bp DNA Homo sapiens genomic DNA, chromosome 21q22.1, D21S22c-ANL region clone B2444F14-T50B8, segment 1/9, complete sequence.
ACCESSION	AP000165
NTID	94827130
VERSION	AP000165.1
KEYWORDS	HTG.
SOURCE	Homo sapiens DNA.
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
1 (bases 1 to 100000)  
Hattori, M., Ishii, K., Toyoda A  
Tavler, T. D.

JOURNAL REFERENCE	Fujiiyama A., Yada T., Totoki Y., Honjo-Seong P., Homo sapiens B90,291bp genomic DNA of 21q22.1 (REGION: D21S26-AMT CLONE RANGE: 8234fb14-550f8) Published Only in Database (1999) In press 2 (bases 1 to 10000)
AUTHORS	Hattori M., Ishikawa K.

Submitted (10-MAY-1999) to the *Journal of the Hong-Seong, P...*

COMMENT

protein sequences which integrated in the original/previous sequences. We determined the boundary between the insertion and genomic sequences experimentally, removed the insertion sequences, reconstituted the present data. The sequencing project is supported by Japan Science Technology Corporation (JST) and The Institute of Physical and Chemical Research (RIKEN).

Location  
1. 100000  
source

Source

BASE COUNT	26797	a	20563	c	21833	g	30807	t
ORIGIN								

Query Match	69.2%	Score 18;	DB 9;	Length 100000;
Best Local Similarity	80.8%;	Pred. No. 1.5e+02;		
Matches 21; Conservative	0;	Mismatch		

QY 1 ATGCTTCGAGCTCGAAGAAGCAAGATG 26  
||||| ||| | ||||| |||||  
Db 67753 ATGCCCCGAGGCTGTAAGCAAGATG 67778

RESULT	8
AP000315	DNA PRI 15-MAY-1999
LOCUS	68109 bp
DEFINITION	Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AWL region, clone:B2344F14, complete sequence.
ACCESSION	AP000315
NID	94835684
VERSION	AP000315.1 GI:4835684
KEYWORDS	HTG.
SOURCE	Homo sapiens DNA, clone:B2344F14. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ORGANISM	
REFERENCE	
AUTHORS	1 (bases 1 to 68109) Toyoda A., Taylor,T.D., Hong-Seog,P., Hattori,M., Ishii,K., Totoki,Y. and Sakaki,Y. Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y. Homo sapiens 68,109bp genomic DNA of 21q22.1 Published Only in Database (1999) In press 2 (bases 1 to 68109) Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y. Direct Submission Submitted (13-MAY-1999) to the DDBJ/EMBL/Genbank databases. Masahito Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC) Kitasato Univ., 1-15-1 Kitasato, Sagaminara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken-go.jp/ tel:81-42-778-9923, Fax:81-42-778-9924) The sequence is a part of the data (ACCESSION NO. AP000165 - AP000173).
COMMENT	The sequencing project is supported by Japan Science Technology Corporation (JST) and The Institute of Physical and Chemical Research (RIKEN). Location/Qualifiers 1..68109 /db_xref="taxon:9606" /chromosome="21" /clone="B2344F14" /map="21q22.1" /seq_id="13825 c 14920 g 21593 t 17771 a 13825 c 14920 g 21593 t
FEATURES	
SOURCE	
BASE COUNT	17771 a 13825 c 14920 g 21593 t
ORIGIN	
Query Match	69.2%; Score 18; DB 9; Length 68109;
Best Local Similarity	80.8%; Pred. No. 1.6e+02;
Matches	21; Conservative 0; Mismatches 5; Indels 0; Gaps 0
OY	1 ATGCCTCGAGGTCCGAAGAACAAGATG 26                        Db 67753 ATGCCCCCAGCGGTAAAGCAAGATG 67778
RESULT	9
AP000316/c	DNA PRI 15-MAY-1999
LOCUS	24518 bp
DEFINITION	Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AWL region, clone:S185, complete sequence.
ACCESSION	AP000316
NID	94835685
VERSION	AP000316.1 GI:4835685
KEYWORDS	HTG.
SOURCE	Homo sapiens DNA, clone:S185. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ORGANISM	
REFERENCE	
AUTHORS	1 (bases 1 to 24518) Toyoda A., Taylor,T.D., Hong-Seog,P., Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., and Sakaki,Y. Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y. Homo sapiens 24,518bp genomic DNA of 21q22.1 Published Only in Database (1999) In press
TITLE	
JOURNAL	

```
misc_feature
1..1334
/clone_id="RPC1-1"
/sequence="match: multiple ESTs; similar to retroviral
sequence"
prim_transcript 577..926
/sequence="match: EST C04699 clone 3MHC3826"
repeat_region
1335..1787
/sequence="LIR7 repeat: matches 1..450 of consensus"
repeat_region
1788..2034
/sequence="MSRD repeat: matches 251..1 of consensus"
repeat_region
2592..2701
/sequence="LIR7 repeat: matches 1..131 of consensus"
repeat_region
2653..2958
/sequence="LIR7 repeat: matches 134..450 of consensus"
repeat_region
3100..3397
/sequence="LIR7 repeat: matches 134..450 of consensus"
repeat_region
3449..3726
/sequence="LIR7 repeat: matches 594..891 of consensus"
repeat_region
4047..4479
/sequence="LIR7 repeat: matches 1..302 of consensus"
repeat_region
4482..4793
/sequence="LIR7 repeat: matches 4902..4467 of consensus"
repeat_region
4799..5337
/sequence="LIR7 repeat: matches 301..2 of consensus"
repeat_region
5333..5398
/sequence="LIR7 repeat: matches 4478..3939 of consensus"
repeat_region
5398..5789
/sequence="LIR7 repeat: matches 4892..4955 of consensus"
repeat_region
5848..6271
/sequence="LIR7 repeat: matches 1..436 of consensus"
repeat_region
6130..6983
/sequence="LIR7 repeat: matches 4958..5376 of consensus"
repeat_region
8198..8333
/sequence="LIR7 repeat: matches 1..904 of consensus"
repeat_region
8748..9180
/sequence="LIR7 repeat: matches 158..293 of consensus;
incomplete repeat"
repeat_region
9904..10090
/sequence="LIR7 repeat: matches 418..2 of consensus"
repeat_region
10412..10519
/sequence="LIR7 repeat: matches 537..350 of consensus"
repeat_region
10809..10910
/sequence="LIR7 repeat: matches 47..156 of consensus"
repeat_region
12188..12480
/sequence="LIR7 repeat: matches 71..172 of consensus"
repeat_region
12724..12905
/sequence="LIR7 repeat: matches 294..1 of consensus"
repeat_region
12801..13124
/sequence="LIR7 repeat: matches 83..262 of consensus"
repeat_region
13201..13485
/sequence="LIR7 repeat: matches 236..8 of consensus"
repeat_region
13598..13894
/sequence="LIR7 repeat: matches 297..12 of consensus"
repeat_region
13909..13974
/sequence="LIR7 repeat: matches 10..344 of consensus"
repeat_region
14136..14336
/sequence="LIR7 repeat: matches 1055..990 of consensus"
repeat_region
14722..14916
/sequence="LIR7 repeat: matches 7..213 of consensus"
prim_transcript 17399..21848
/sequence="match: multiple ESTs; similar to S37431 laminin
receptor; match: C05508 W95930 F01047 T59359 N89876;
match: T59387 W57792 H61800 D19641 H81305; match: H30870
F17735 N25161 N57410 H49137; match: W01270 R62252 N46183
W10275 W03462; match: R44437 W73502 N67862 T53483 W19436;
match: T94056 N86655 W52358 W19382 N89493; match: T53496
R97292"
repeat_region
21987..22358
/sequence="THEIC repeat: matches 371..3 of consensus"
repeat_region
22361..23983
/sequence="THEIC repeat: matches 1580..4 of
consensus"
repeat_region
24005..24392
/sequence="THEIC repeat: matches 371..1 of consensus"
repeat_region
24389..24916
/sequence="LIR7 repeat: matches 2110..2643 of consensus"
repeat_region
24940..25215
/sequence="LIR7 repeat: matches 302..27 of consensus;
incomplete repeat"
repeat_region
25396..27397
/sequence="LIR7 repeat: matches 2874..4908 of consensus"
repeat_region
27408..27453
/sequence="LIR7 repeat: matches 906..951 of consensus"
repeat_region
27729..28146
/sequence="LIR7 repeat: matches 1043..657 of consensus"
repeat_region
29126..29494
/sequence="LIR7 repeat: matches 1043..657 of consensus"
repeat_region
29492..29603
/sequence="LIR7 repeat: matches 594..670 of consensus"
repeat_region
29492..29565
/sequence="LIR7 repeat: matches 594..670 of consensus"
repeat_region
30350..30589
/sequence="LIR7 repeat: matches 4..262 of consensus"
repeat_region
31122..31249
/sequence="LIR7 repeat: matches 4..262 of consensus"
repeat_region
31643..31838
/sequence="LIR7 repeat: matches 1055..861 of consensus"
repeat_region
33626..33705
/sequence="LIR7 repeat: matches 144..56 of consensus"
repeat_region
35804..36169
/sequence="LIR7 repeat: matches 911..535 of consensus"
repeat_region
36166..36380
/sequence="LIR7 repeat: matches 416..196 of consensus"
repeat_region
38150..38235
/sequence="LIR7 repeat: matches 745..664 of consensus"
repeat_region
41350..41413
/sequence="LIR7 repeat: matches 261..20 of consensus"
repeat_region
43481..43872
/sequence="LIR7 repeat: matches 406..4 of consensus"
repeat_region
46750..46904
/sequence="LIR7 repeat: matches 1051..892 of consensus"
repeat_region
47485..48066
/sequence="LIR7 repeat: matches 311..903 of consensus"
repeat_region
51925..52166
/sequence="LIR7 repeat: matches 261..20 of consensus"
repeat_region
52167..52470
/sequence="LIR7 repeat: matches 1..301 of consensus"
repeat_region
52689..53086
/sequence="LIR7 repeat: matches 1..405 of consensus"
repeat_region
53330..53565
/sequence="LIR7 repeat: matches 236..2 of consensus;
incomplete repeat"
repeat_region
55341..55616
/sequence="LIR7 repeat: matches 293..1 of consensus"
repeat_region
56208..56527
/sequence="LIR7 repeat: matches 1..300 of consensus"
repeat_region
56851..56928
/sequence="LIR7 repeat: matches 374..293 of consensus"
repeat_region
56925..57111
/sequence="LIR7 repeat: matches 211..5 of consensus"
repeat_region
59278..59437
/sequence="LIR7 repeat: matches 897..1054 of consensus"
repeat_region
59523..59730
/sequence="LIR7 repeat: matches 26..245 of consensus"
repeat_region
61233..61451
/sequence="LIR7 repeat: matches 4..189 of consensus"
repeat_region
61617..61920
/sequence="LIR7 repeat: matches 303..1 of consensus"
repeat_region
66259..66619
/sequence="LIR7 repeat: matches 1681..2041 of consensus"
repeat_region
67157..67244
/sequence="LIR7 repeat: matches 192..104 of consensus"
BASE COUNT 22764 a 12519 c 12456 g 21274 t
ORIGIN
```





Thu Sep 23 10:37:29 1999

us-08-671-757-1.rge

Page 12

FEATURES	
Location/Qualifiers	* be preserved.
1 135437	

**source**

```
1. .135437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="796E4"
```

BASE COUNT					
ORIGIN	37381	a	28828	c	28389 g 39239 t 1600 others

[illegible]

Query Match	69.2%	Score 18;	DB 34;	Length 135437;
Best Local Similarity	80.8%	Pred No. 1.5e+02;		
Matches 21;	Conservative	0;	Mismatches 5;	Indels 0;
			Gaps	0;

Query Match	69.2%;	Score 18;	DB 34;	Length 1334377;
Best Local Similarity	80.8%;	Pred. No. 1.5e+02;		
Matches 21;	Conservative	0;	Mismatches 5;	Indels 0;

0;	Mismatches	5;	Indels	0;	Gaps	0;
----	------------	----	--------	----	------	----

QY 1 ATGCTCGAGTTCGAAAGCAGATG 26  
||||| | | | |||||  
Db 21136 ATGCTTGTGTGTCAAGCAGATG 21161

Search completed: September 22, 1999, 17:01:57  
Job time: 496 sec

Search completed: September 22, 1999, 17:01:57  
Job time: 496 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: September 22, 1999, 16:53:41 ; Search time 51.18 Seconds

(without alignments)  
127.100 Million cell updates/sec

Title: US-08-671-757-1

Perfect score: 26  
Sequence: 1 ATGCTCGAGTGCAGAAACGACAGATG 26

Scoring table: IDENTITY\_NNC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	26	100.0	26	1	T62334	Helicobacter pylori
2	18.2	70.0	110000	1	V30458-1	Continuation (2 of
3	18.2	70.0	110000	1	V30459-1	Continuation (2 of
4	17.6	67.7	3960	1	O48732	CT gene, Recombina
5	17.2	66.2	98	1	O81918	Interferon gamma b
6	16.4	63.1	876	1	T61605	Humicola lanuginos
7	16.4	63.1	876	1	T61606	Humicola lanuginos
8	16.4	63.1	876	1	T61593	Humicola lanuginos
9	16.4	63.1	876	1	T61594	Humicola lanuginos
10	16.4	63.1	31	1	V72952	Rat Munc13-1 PCR p
11	16.4	63.1	1	1	X20600	Polynucleotide seq
12	16.2	62.3	2163	1	T95753	Arabidopsis SCAREC
13	16	61.5	3649	1	O04476	Sequence derived b
14	16	61.5	1657	1	N81265	Gene encoding Chla
15	16	61.5	99	1	N00005	Sequence encoding
16	16	61.5	2518	1	T31256	Hydantoinease codin
17	16	61.5	3190	1	T31257	Continuation (6 of
18	16	61.5	80073	1	T58840-5	Glutaryl-tRNA(Gln)
19	16	61.5	1467	1	X20935	Polynucleotide seq
20	16	61.5	240	1	X17999	DNA encoding a RNA
21	16	61.5	1467	1	X16047	Human kidney amino
22	15.8	60.8	37	1	V19310	Rat p26 gene RACE
23	15.8	60.8	50000	1	X23517	Human kidney amino
24	15.6	60.0	419	1	O03280	Portion of plasmid
25	15.6	60.0	1714	1	N90718	Oct-2 clone encodi
26	15.6	60.0	1847	1	N90719	Pass-3 clone encod
27	15.6	60.0	5393	1	O34611	Plasmid pR3. Immun
28	15.6	60.0	4641	1	O40294	Sequence of plasmid
29	15.6	60.0	6811	1	O41004	Sequence of plasmid
30	15.6	60.0	5472	1	O41005	Sequence of plasmid
31	15.6	60.0	1133	1	O40246	Sequence of a DNA
32	15.6	60.0	1133	1	O40247	Sequence of a DNA
33	15.6	60.0	1209	1	O40264	Sequence inserted
34	15.6	60.0	4145	1	O40279	Sequence of clone
35	15.6	60.0	4277	1	O40280	Sequence of clone
36	15.6	60.0	6926	1	O40281	Sequence of clone
37	15.6	60.0	32	1	O93242	Plasmid todic12BA
38	15.6	60.0	9020	1	T07652	PSG11 FlpV N. Reco
39	15.6	60.0	8710	1	O89650	PSG11 FlpV El. Rec
40	15.6	60.0	613	1	T01289	Rhizobium species
41	15.6	60.0	900	1	T01289	Banana polyphenol
42	15.6	60.0	28984	1	V15826	Genomic DNA for in
43	15.6	60.0	4185	1	V43563	Human blue-light p

## ALIGNMENTS

44 15.6 60.0 9767 1 X13067  
45 15.6 60.0 6560 1 X26118

Enterococcus faeca  
DNA sequence of FP

## RESULT 1

ID T62334  
AC T62334;  
DE 23-SEP-1997 (first entry)  
DE Helicobacter pylori flbA gene PCR primer, Olf1A-1.  
KW Flagellum; motility; monoclonal antibody; vaccine; infection; ulcer;  
KW ulcerative colitis; chronic gastritis; duodenal; immunise;  
KW prevention; hook protein; attenuated urease; primer; probe; PCR;  
KW Helicobacter pylori.  
OS EP-752473-A2.  
PN 08-JAN-1997.  
PD 28-JUN-1996; 401445.  
PR 04-JUL-1995; RR-008068.  
PA (INRM) INST NAT SANTE & RECH MEDICALE.  
PA (INSP) INST PASTEUR.  
PI Labigne A, Suerbaum S;  
DR WPI; 97-067450/07.  
PT Helicobacter pylori flbA gene and related protein - regulates  
PT biosynthesis of flagella, useful in compns. and vaccines for  
PS Claim 1; Page 38; 58pp; French.  
CC T62334 and T62335 are PCR primers/probes used for the isolation of the  
CC Helicobacter flbA gene. The flbA gene regulates the biosynthesis of  
CC H. pylori flagellum proteins. The flbA gene may be mutated to produce  
CC a strain of H. pylori having an attenuated phenotype. H. pylori  
CC defective (or absent) hook protein. Alternatively the flbA protein can  
CC be used to raise monoclonal antibodies or polyclonal serum for  
CC detecting H. pylori infections in vitro or to prepare vaccines against  
CC H. pylori infections. Vaccines especially also contain H. pylori antigens  
CC encoded by the urea, ureb, urec or ured genes.  
SQ Sequence 26 BP; 9 A; 5 C; 8 G; 4 T;

Query Match 100.0%; Score 26; DB 1; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGCTCGAGTGCAGAAACGACAGATG 26  
DB 1 ATGCTCGAGTGCAGAAACGACAGATG 26

RESULT 2  
V30458-1/c  
Continuation (2 of 6) of V30458 from base 100001 (Rhizobium species plasmid pMGR234a.  
WP Sequence split into 6 fragments LOCUS V30458 Accession V30458

WP Fragment Name Begin End  
WP V30458-0 1 110000  
WP V30458-1 100001 210000  
WP V30458-2 200001 310000  
WP V30458-3 300001 410000  
WP V30458-4 400001 510000  
WP V30458-5 500001 534720

Query Match 70.0%; Score 18.2; DB 1; Length 110000;  
Best Local Similarity 87.0%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 TGCCTCGAGTGCAGAAACGACAGA 24  
DB 72074 TGCCTCGAGTGCAGAAACGACAGA 72052





PT Preparing polypeptide variants with improved functional properties -  
PT by in vivo recombination between opened plasmid and homologous DNA,  
PT to produce e.g. enzymes with improved washing and dishwashing  
PT properties  
PS Example 1; Fig 1; 68pp; English.  
CC A DNA sequence (T61605) comprises the Humicola lanuginosa wild-type  
CC lipase gene in Saccharomyces cerevisiae expression plasmid pJS026.  
CC It encodes a polypeptide (W13560) of 291 amino acid residues.  
CC Plasmid pJS026 has been used to demonstrate an improved method of  
CC preparing positive polypeptide variants. This involves shuffling  
CC homologous DNA sequences in an iterative in vivo recombination  
CC system using a eukaryotic cell (such as yeast) as a recombination  
CC host cell. In an example, the lipase gene from pJS026 was  
CC transformed into S. cerevisiae YNG318 together with a plasmid,  
CC pJS037 (see also T61606), which carries a synthetic H. lanuginosa  
CC lipase gene and which had been opened at approx. the middle of the  
CC lipase coding sequence by digestion with NruI, PstI, and NruI and  
CC PstI. 50% of transformants contained recombinant DNA sequences.  
SQ Sequence 876 BP; 208 A; 229 C; 225 G; 214 T;

Query Match 63.1%; Score 16.4; DB 1; Length 876;  
Best Local Similarity 76.9%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 1 ATGCTCGAGGTGCAAGCAAGATG 26  
186 ATGCCCGAGGTAGAGAGCGGATG 211

RESULT 7  
T61606 ID' T61606 standard; DNA; 876 BP.  
AC T61606;  
DE 21-OCT-1997 (first entry)  
DE Humicola lanuginosa lipase synthetic gene in pJS037.  
KW Lipase; polypeptide variant; in vivo recombination; shuffling;  
KW Saccharomyces cerevisiae; detergent; pJS037; ss.  
OS Synthetic.  
PN WO9707206-A1.  
PD 27-FEB-1997.  
PF 12-AUG-1996; DK0344.  
PR 20-SEP-1995; DK-001047.  
PR 11-AUG-1995; DK-000907.  
PA (NOVO) NOVO-NORDISK AS.  
PI Okkels JS;  
DR WPI: 97-165290/15.  
DR P-PSDB; W13560.  
PT Preparing polypeptide variants with improved functional properties -  
PT by in vivo recombination between opened plasmid and homologous DNA,  
PT to produce e.g. enzymes with improved washing and dishwashing  
PT properties  
PS Example 1; Fig 2; 68pp; English.  
CC A DNA sequence (T61605) comprises a Humicola lanuginosa synthetic  
CC lipase gene in Saccharomyces cerevisiae expression plasmid pJS037.  
CC It contains 12 additional restriction sites not found in the wild-  
CC type gene (see also T61605). Plasmid pJS037 has been used to  
CC demonstrate an improved method for preparing positive polypeptide  
CC variants. This involves shuffling homologous DNA sequences in an  
CC iterative in vivo recombination system using a eukaryotic cell  
CC (such as yeast) as a recombination host cell. In an example, in  
CC vivo recombination of H. lanuginosa lipase variants was performed  
CC using NruI-cut pJS037 and S. cerevisiae YNG318 as host cell. 2  
CC transformants (see also W13561-62) were obt'd. that showed improved  
CC wash performance.  
SQ Sequence 876 BP; 205 A; 231 C; 224 G; 216 T;

Query Match 63.1%; Score 16.4; DB 1; Length 876;  
Best Local Similarity 76.9%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 1 ATGCTCGAGGTGCAAGCAAGATG 26

Db 186 ATGCCCGAGGTAGAGAGCGGATG 211

RESULT 8  
T61593 ID' T61593 standard; DNA; 876 BP.  
AC T61593;  
DE 21-OCT-1997 (first entry)  
DE Humicola lanuginosa lipase gene in vector pJS026.  
KW Lipase; polypeptide variant; in vivo recombination; shuffling;  
KW Saccharomyces cerevisiae; detergent; pJS026; ss.  
OS Humicola lanuginosa DSM 4109.  
PN WO9707205-A1.  
PD 27-FEB-1997.  
PF 12-AUG-1996; DK0343.  
PR 20-SEP-1995; DK-001047.  
PR 11-AUG-1995; DK-000907.  
PA (NOVO) NOVO-NORDISK AS.  
PI Okkels JS;  
DR WPI: 97-165289/15.  
DR P-PSDB; W13556.  
PT Preparing polypeptide variants with improved functional properties -  
PT by in vivo recombination between opened plasmid and homologous DNA,  
PT to produce e.g. enzymes with improved washing and dishwashing  
PT properties  
PS Example 1; Fig 1; 68pp; English.  
CC A DNA sequence (T61593) comprises the Humicola lanuginosa wild-type  
CC lipase gene in Saccharomyces cerevisiae expression plasmid pJS026.  
CC It encodes a polypeptide (W13556) of 291 amino acid residues.  
CC Plasmid pJS026 has been used to demonstrate an improved method of  
CC preparing positive polypeptide variants. This involves shuffling  
CC homologous DNA sequences in an iterative in vivo recombination  
CC system using a eukaryotic cell (such as yeast) as a recombination  
CC host cell. In an example, the lipase gene from pJS026 was  
CC transformed into S. cerevisiae YNG318 together with a plasmid,  
CC pJS037 (see also T61594), which carries a synthetic H. lanuginosa  
CC lipase gene and which had been opened at approx. the middle of the  
CC lipase coding sequence by digestion with NruI, PstI, and NruI and  
CC PstI. 50% of transformants contained recombinant DNA sequences.  
SQ Sequence 876 BP; 208 A; 229 C; 225 G; 214 T;

Query Match 63.1%; Score 16.4; DB 1; Length 876;  
Best Local Similarity 76.9%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 186 ATGCCCGAGGTAGAGAGCGGATG 211

RESULT 9  
T61594 ID' T61594 standard; DNA; 876 BP.  
AC T61594;  
DE 21-OCT-1997 (first entry)  
DE Humicola lanuginosa lipase synthetic gene in pJS037.  
KW Lipase; polypeptide variant; in vivo recombination; shuffling;  
KW Saccharomyces cerevisiae; detergent; pJS037; ss.  
OS Synthetic.  
PN WO9707205-A1.  
PD 27-FEB-1997.  
PF 12-AUG-1996; DK0343.  
PR 20-SEP-1995; DK-001047.  
PR 11-AUG-1995; DK-000907.  
PA (NOVO) NOVO-NORDISK AS.  
PI Okkels JS;  
DR WPI: 97-165289/15.  
DR P-PSDB; W13556.  
PT Preparing polypeptide variants with improved functional properties -  
PT by in vivo recombination between opened plasmid and homologous DNA,  
PT to produce e.g. enzymes with improved washing and dishwashing

PT Properties  
PS Example 1; Fig 2: 68pp: English.  
CC A DNA sequence (T61594) comprises a Humicola lanuginosa synthetic  
CC lipase gene in Saccharomyces cerevisiae expression plasmid pJS037.  
CC It contains 12 additional restriction sites not found in the wild-  
CC type gene (see also T61593). Plasmid pJS037 has been used to  
CC demonstrate an improved method for preparing positive polypeptide  
CC variants. This involves shuffling homologous DNA sequences in an  
CC iterative in vivo recombination system using a eukaryotic cell  
CC (such as yeast) as a recombination host cell. In an example,  
CC in vivo recombination of H. lanuginosa lipase variants was  
CC performed using NuiI-cut pJS037 and S. cerevisiae YN318 as  
CC host cell. 2 transformants (see also W13557-58) were obtd. that  
CC showed improved wash performance.  
SQ Sequence 876 BP; 205 A; 231 C; 224 G; 216 T;

Query Match 63.1%; Score 16.4; DB 1; Length 876;  
Best Local Similarity 76.9%; Pred. No. 74;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
OY 1 ATGCTCGAGTCAAGACAGATG 26  
Db 186 ATGCTCGAGTCAAGACAGATG 211  
||||| ||||| ||||| ||||| |||||

RESULT 10  
V72952  
ID V72952 standard; DNA: 31 BP.  
AC V72952;  
DT 04-MAR-1999 (first entry)  
DE Rat Munc13-1 PCR primer SEQ ID NO:11.  
KW Munc13; Doc2-alpha; interacting domain; screening; agonist; antagonist;  
KW calcium ion dependent secretion inhibitor; neurotransmitter; hormone;  
KW fusion protein; nervous disease; PCR primer; ss.  
OS Synthetic.  
OS Rattus sp.  
PN J10313866-A.  
PD 02-DEC-1998.  
PF 15-MAY-1997; 126118.  
PR 15-MAY-1997; JP-126118.  
PT (SHITO) SHIONOGI & CO LTD.  
DR WPI; 99-074148/07.  
PT Screening for agonists or antagonists of binding between Doc2-alpha  
PT and Munc13 - used to treat diseases of the nervous system  
PS Example 3; Page 27; 33pp: Japanese.  
CC The present invention describes a method of screening for agonists or  
CC antagonists of the binding between Doc2-alpha and Munc13. The method  
CC comprises reacting Doc2-alpha or its homologue with Munc13 or its  
CC homologue optionally in the presence of a test substance and selecting  
CC the substances which increase or decrease binding. Also described are:  
CC (1) an agonist or antagonist of the binding between Doc2-alpha and  
CC Munc13 selected by the above method; (2) a vector expressing Doc2-alpha  
CC or its homologue used for inhibiting Ca ion-dependent secretion of a  
CC neurotransmitter or hormone; (3) a vector expressing Munc13 or its  
CC homologue used for inhibiting Ca ion-dependent secretion of a  
CC neurotransmitter or hormone; (4) a fusion protein between Doc2-alpha or  
CC its homologue and a carrier protein; (5) a fusion protein containing  
CC or its homologue and a carrier protein; (6) a polypeptide containing  
CC amino acids 13-37 of the sequence of Doc2-alpha, which binds with Munc13  
CC and comprises at most 90 amino acids; and (7) a polypeptide containing  
CC amino acids 851-1461 of the sequence of Munc13, which binds with Doc2-  
CC alpha and comprises at most 904 amino acids. The agonist or antagonist  
CC can be used to treat diseases of the nervous system. The present  
CC sequence represents a PCR primer for rat Munc13-1.  
SQ Sequence 31 BP; 10 A; 5 C; 11 G; 5 T;

Query Match 63.1%; Score 16.4; DB 1; Length 31;  
Best Local Similarity 76.9%; Pred. No. 52;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1 ATGCTCGAGTCAAGACAGATG 26

Db 6 ATGCTCGAGTCAAGACAGATG 31  
||||| ||||| ||||| ||||| |||||

RESULT 11  
X20600/C  
ID X20600 standard; DNA: 22191 BP.  
AC X20600;  
DT 05-MAY-1999 (first entry)  
DE Polynucleotide sequence from the genome of Treponema pallidum.  
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;  
KW enzyme production; ds.  
OS Treponema pallidum.  
PN W09859034-A2.  
PD 30-DEC-1998.  
PF 23-JUN-1998; U13041.  
PR 24-JUN-1997; US-050667.  
PT (HUMA-) HUMAN GENOME SCI INC.  
PT Fraser CM;  
PT WPI; 99-081273/07.  
PT New isolated Treponema pallidum nucleic acids - used to develop  
PT products for the detection, diagnosis, characterisation, prevention  
PT and therapy of T. pallidum infections, particularly syphilis  
PS Claim 1; Page 673-685; 1150pp: English.  
CC X20300-21243 represent polynucleotide sequences from the genome of  
CC Treponema pallidum. The sequences can be used for detection,  
CC diagnosis, characterisation, prevention and therapy for T. pallidum  
CC infections, particularly syphilis. They can also be used for detecting  
CC diseases related to Borrelia infections in animals, and for the  
CC production of biosynthetic products such as enzymes.  
SQ Sequence 22191 BP; 4744 A; 5303 C; 6215 G; 5916 T;

Query Match 63.1%; Score 16.4; DB 1; Length 22191;  
Best Local Similarity 76.9%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 ATGCTCGAGTCAAGACAGATG 26  
Db 15713 AGCCTCGAGTCAAGACAGATG 15688  
||||| ||||| ||||| ||||| |||||

RESULT 12  
T95753/C  
ID T95753 standard; DNA: 2163 BP.  
AC T95753;  
DT 11-MAY-1998 (first entry)  
DE Arabidopsis SCARECROW gene.  
KW SCARECROW; SCR gene; transgenic plant; root; gravitropism;  
KW crop improvement; ds.  
OS Arabidopsis thaliana.  
FH Key  
FT Location/Qualifiers  
FT CDS  
FT 49..2125  
FT /\*tag= a  
FT /note= "contains an intron"  
FT 49..1580  
FT /\*tag= b  
FT /number= 1  
FT intron 1581..1695  
FT /\*tag= c  
FT /number= 1  
FT exon 1696..2125  
FT /\*tag= d  
FT /number= 2  
PN W09741152-A1.  
PD 06-NOV-1997.  
PF 25-APR-1997; U07022.  
PR 24-APR-1997; US-842445.  
PR 26-APR-1996; US-638617.  
PA (UNYK) UNIV NEW YORK STATE.  
PI Bentley PN, Dilaurenzio L, Helariutta Y, Malamy JE,  
PI Pysh L, Wysocka-Diller J;  
DR WPI; 97-549683/50.

DR P-PSDB; W38178.  
 PT DNA encoding Arabidopsis SCARECROW protein - useful to modify plant  
 PT cell division and therefore alter root development, or alter plant  
 PS stem or hypocotyl gravitropism  
 CC Claim 6: Page 90-92; 221pp; English.  
 CC This sequence comprises the Arabidopsis SCARECROW (SCR gene)  
 CC genomic region. It was obtained following isolation of an scr-1  
 CC mutant from a T-DNA transformed Arabidopsis seedling that had  
 CC greatly reduced root length. Plant DNA flanking the insertion site  
 CC was isolated from scr-1 by plasmid rescue, and used to isolate the  
 CC wild-type genomic DNA. The SCR gene encodes a 653-amino acid  
 CC protein (see W38178). It is expressed specifically in embryo root  
 CC progenitor tissue and in certain root and stem tissues. It  
 CC regulates a specific asymmetric division, and controls gravitropic  
 CC response in aerial structures and root formation. Nucleic acid  
 CC W38178-201), SCR proteins lacking 1 to 4 of MOTIFS I to VI, or  
 CC SCR MOTIF I, II, III, IV, V or VI are claimed. Transgenic plants  
 CC can be engineered to overexpress the SCR protein, so that cell  
 CC division is increased in roots, resulting in thicker root  
 CC development, while a plant with an altered stem or hypocotyl  
 CC gravitropism is less susceptible to lodging. Plants that contain  
 CC an antisense molecule that suppresses the expression of endogenous  
 CC SCR gene product have thinner root development. A gene of interest  
 CC can be placed under control of a SCR promoter and expressed in a  
 CC plant to confer herbicide, salt, pathogen or insect resistance, or  
 CC when expressed in stems to increase starch, lignin or cellulose  
 CC biosynthesis (all claimed).  
 SQ Sequence 2163 BP; 571 A; 546 C; 439 G; 607 T;

Query Match  
 Best Local Similarity 62.3%; Score 16.2; DB 1; Length 2163;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 CCTCGAGCTGGAAGCAAGA 24  
 Db 1328 CCTCGAGCTGGAAGCAAGA 1308

RESULT 13  
 ID 004476  
 AC 004476 standard; cDNA; 3649 BP.  
 DT 05-OCT-1990 (first entry)  
 DE Sequence derived by reverse transcription of RNA1 of AMV.  
 KW Alfalfa mosaic virus; AMV; cryptic infection; probe; ss.  
 OS Nicotiana tabacum L. "Xanthi-nc"  
 FH Key Location/Qualifiers  
 FT cds 100..3486  
 FT /\*tag= a  
 FT /\*note="protein encoded by RNA1"  
 PD US4921802-A.  
 PD 01-MAY-1990.  
 PF 13-JAN-1988; 144692.  
 PR 05-MAY-1983; US-492582.  
 PR 13-JAN-1988; US-144692.  
 PA (PRO) Pioneer Hi-Bred International Ltd.  
 PI Hall TC, Loesch-Fries SL, Jarvis NP, Barker RF;  
 DR WPI; 90-171297/22.  
 DR P-PSDB; R05107.  
 PT Recombinant cDNA plasmids derived from alfalfa mosaic virus RNA -  
 PT for detection of cryptic infection by alfalfa mosaic virus.  
 PS Disclosure: P; English.  
 CC The sequence is that of the bottom (B) - or "RNA1" genome fragment  
 CC of AMV which is necessary to initiate infection. The sequence can  
 CC be used as a probe to detect cryptic infection of plants by AMV.  
 CC See also 004477 and 78.  
 SQ Sequence 3649 BP; 1039 A; 744 C; 808 G; 1058 T;

Query Match  
 Best Local Similarity 61.5%; Score 16; DB 1; Length 3649;  
 Matches 79.2%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 3 GCCTCGAGGCGGAAAGCAAGATG 26  
 Db 182 GACCGCTACTCGAAGCAAGCTG 205

RESULT 14  
 ID N81265/c  
 AC N81265; standard; DNA; 1657 BP.  
 DT 24-OCT-1990 (first entry)  
 DE Gene encoding Chlamydia elementary body protein, chlamydia.  
 KW Chlamydia; elementary body; binding protein; vaccine; ss.  
 OS Chlamydia trachomatis serovar L2 (L2/Bu/434).  
 FH Key Location/Qualifiers  
 FT rbs 146..149  
 FT cds /\*tag= a  
 FT /\*tag= b  
 FT /\*tag= c  
 FT signal\_peptide prod-chlamydia  
 FT note="18 kD binding protein"  
 FT /\*tag= c  
 PD EP-293079-A.  
 PD 30-NOV-1988.  
 PF 21-APR-1988; 303584.  
 PR 24-APR-1987; GB-009746.  
 PA (THRE-) Alberta Ltd.  
 PI Weinman WM, Kaul R;  
 DR WPI; 88-39641/48.  
 DR P-PSDB; P80763.  
 PT New antigenic polypeptide cpds. from Chlamydia trachomatis -  
 PT useful in vaccines, and new DNA sequences encoding them.  
 PS Claim 2: Fig 3b; 19pp; English.  
 CC The 18 kD product of the gene is an antigen involved in attachment  
 CC of Chlamydia elementary body to the host cell. Antibodies raised  
 CC against the protein can be used for passive immunisation against  
 CC Chlamydia infections.  
 SQ Sequence 1657 BP; 477 A; 319 C; 389 G; 472 T;

Query Match  
 Best Local Similarity 61.5%; Score 16; DB 1; Length 1657;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TGCCGAGCTGGAAGCAAGAT 25  
 Db 901 TTCCGAGGCGGCAAGCAAGATAT 878

RESULT 15  
 ID N00005/c  
 AC N00005; standard; cDNA; 99 BP.  
 DT 14-OCT-1992 (first entry)  
 DE Sequence encoding N-terminal of influenza haemagglutinin gene.  
 KW Vaccine; influenza gene; haemagglutinin; antigen; ss.  
 OS Fowl pest virus (Rostock strain).  
 FH Key Location/Qualifiers  
 FT cds 22..99  
 FT /\*tag= a  
 PD BE-882545-A.  
 PD 30-SEP-1980.  
 PF 31-MAR-1980; 882545.  
 PR 02-APR-1979; GB-011487.  
 PA (SEAR) SEARLE G D & CO.  
 DR WPI; 80-73458C/42.  
 DR P-PSDB; P00009.  
 PT Synthetic influenza gene prodn. - from viral RNA by inverse  
 PT transcription then converting DNA to double helix coding for  
 PT antigenic proteins when incorporated in plasmid(s)  
 PS Disclosure; Fig 5; 23pp; French.

CC Viral RNA (vRNA) was isolated from fowl pest virus (Pestock Strain)  
CC subjected to polyadenylation with ATP/poly(A) polymerase then the  
CC corresp. DNA was synthesised using inverse transcriptase (IT).  
CC N00004 shows the structure of vRNA for influenza haemagglutinin,  
CC indicating the posns. of the start and stop codons. N00005 and  
CC N00006 give the sequences of cDNA corresp. to bps 1-98 and 1028-1129  
CC of N00004 respectively.  
SQ Sequence 99 BP; 32 A; 24 C; 19 G; 24 T;

Query Match	61.5%;	Score 16;	DB 1;	Length 99;
Best Local Similarity	79.2%;	Pred. No. 90;		
Matches 19;	Conservative	0;	Mismatches	5;
			Indels	0;
			Gaps	0;
QY	2	TGCCTCGAGGTGCAAAACAGAT	25	
Db	57	TGCCCAAGAGGCGAAAAACAGAT	34	

Search completed: September 22, 1999, 17:00:01  
Job time: 380 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 1999, 16:53:40 ; Search time 307.21 seconds

(without alignments)  
166.941 Million cell updates/sec

Title: US-08-671-757-1

Sequence: 1 ATGCTCGAGTGCAGAAACCAAGATG 26

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:  
1: em\_est1:  
2: em\_est2:  
3: em\_est3:  
4: em\_est4:  
5: em\_est5:  
6: em\_est6:  
7: em\_est7:  
8: em\_est8:  
9: em\_est9:  
10: em\_est10:  
11: em\_est11:  
12: em\_est12:  
13: em\_est13:  
14: em\_est14:  
15: em\_est15:  
16: em\_est16:  
17: em\_est17:  
18: em\_est18:  
19: em\_est19:  
20: em\_est20:  
21: em\_est21:  
22: em\_est22:  
23: em\_est23:  
24: em\_est24:  
25: em\_est25:  
26: em\_est26:  
27: em\_est27:  
28: em\_est28:  
29: em\_est29:  
30: em\_est30:  
31: em\_est31:  
32: em\_est32:  
33: em\_est33:  
34: em\_est34:  
35: em\_est35:  
36: em\_est36:  
37: em\_est37:  
38: em\_est38:  
39: em\_est39:  
40: em\_est40:  
41: em\_est41:  
42: em\_est42:  
43: em\_est43:  
44: em\_est44:  
45: em\_est45:  
46: em\_est46:  
47: em\_est47:  
48: em\_est48:  
49: em\_est49:  
50: em\_est50:  
51: em\_est51:  
52: em\_est52:  
53: em\_est53:

54: em\_est22:  
55: em\_est23:  
56: em\_est24:  
57: em\_est25:  
58: em\_est26:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.6	71.5	435	22	R23925	R23925 yh28g10.r1
C 2	18	69.2	203	39	AA881975	AA881975 vx31g10.r
C 3	17.6	67.7	524	28	AA116559	AA116559 mp99c11.r
C 4	17.6	67.7	432	30	AA257275	AA257275 MBAFCZ17
C 5	17.6	67.7	459	30	AA266482	AA266482 va05d11.r
C 6	17.6	67.7	310	30	AA271856	AA271856 vb73h12.r
C 7	17.6	67.7	432	40	AA981298	AA981298 vx60h11.r
C 8	17.6	67.7	409	47	AT510391	AT510391 mp99c11.y
C 9	17.6	67.7	493	48	AT608046	AT608046 va05d11.y
C 10	17.6	67.7	347	49	AT661202	AT661202 va05d11.x
C 11	17.2	66.2	361	25	N43298	N43298 SW31CA768SK
C 12	17.2	66.2	590	34	AA503372	AA503372 ne44e12.s
C 13	17.2	66.2	612	41	AT047861	AT047861 ud66f06.y
C 14	17.2	66.2	423	23	R85938	R85938 yo37g06.s1
C 15	17.2	66.2	443	24	H80272	H80272 yu59f01.s1
C 16	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 17	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 18	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 19	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 20	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 21	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 22	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 23	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 24	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 25	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 26	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 27	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 28	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 29	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 30	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 31	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 32	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 33	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 34	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 35	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 36	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 37	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 38	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 39	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 40	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 41	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 42	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 43	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 44	17.2	66.2	445	24	H80286	H80286 yu59f01.s1
C 45	17.2	66.2	445	24	H80286	H80286 yu59f01.s1

## ALIGNMENTS

RESULT 1  
R23925/c R23925 435 bp mRNA  
LOCUS yh28g10.r1 Scores Placenta Nb2HP Homo sapiens CDNA clone  
DEFINITION IMAGE:131106 5', mRNA sequence.  
ACCESSION R23925  
NID 9778813  
VERSION R23925.1 GI:778813



Unpublished (1996)  
On Sep 12, 1996 this sequence replaced ri:13326889

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LINT; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information  
MGT:352012

Location/Qualifiers  
1..524

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="7 q35-q36:4p16.3; 3; X"
/clone="IMAGE:577364"
/clone_lib="Scares 2NDMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="Dh10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer (5'
GTTCACATCGTCAAGTGGAGCGCGCGGTTTTTTTTTTTTTTTTTTT
3'); double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bettina Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Felina Bonaldo."
93 c 108 g 136 t 1 others

```

186 a	93 c	108 g	136 t	1 others
-------	------	-------	-------	----------

Query Match	67.7%	Score 17.6	DB 28	Length 524
Best Local Similarity	83.3%	Pred. NO. 1.7e+02		
Matches 20	Conservative 0	Mismatches 4	Indels 0	Gaps 0
3	GCCTGAGAGTGC	AAAGCAAGATG	26	
325	GCCTGAGAGATG	AAAGCAAGATG	348	

AA257275 432 bp mRNA EST 17-MAR-1997  
 ABACFC21A7T3 Brugia malayi adult female cDNA (SAM96MLW-BmaF) Brugia  
 malayi CDNA clone AFc21A7 5', mRNA sequence.  
 AA257275  
 11893005  
 11893005 GI:1893005  
 EST.  
 Brugia malayi.  
 Brugia malayi

REFERENCE (bases 1 to 432)  
 AUTHORS Blaxter M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and  
 Jones, S. J.  
 TITLE Genes expressed in adult female *Brugia malayi*  
 JOURNAL Unpublished (1996)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395433.  
 Contact: Blaxter ML  
 Institute of Cell, Animal and Population Biology

Location/Qualifiers  
1. .432

University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 1JH, UK.  
3J7F, UK.  
Tel.: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaizot@ed.ac.uk  
The ABI trace of this sequence can be viewed at  
<http://www.sanger.ac.uk/bugia/AFC/MBAF21A73.html>  
Seq primer: 13.

129 a	75 c	101 g	127 t
-------	------	-------	-------

Query Match	67.7%	Score 17.6	DB 30	Length 432
Best Local Similarity	83.3%	Pred. No 1.6e+02		
Matches 20	Conservative 0	Mismatches 4	Indels 0	Gaps 0
QY	2	TGCGTCGAGCTCGAAGACAGAT	25	
Db	323	TGCGTCGAGCTCGAAGACAGAT	300	

AA268482	459 bp	mRNA	EST	21-MAR-1997
U050511.r1	Soares mouse	lymph node	MDMLN	Mus musculus
U050511	5', mRNA sequence.			
AA268482				
U1905226				
AA268482.1	GI:1905226			
EST.				
mouse mouse.				
mus musculus				

**TITLE** The WashU-HHMI Mouse EST Project  
**JOURNAL** Unpublished (1996)  
**COMMENT** On Sep 12, 1996 this sequence version replaced gi:1404853

**Contact:** Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mousestewatson.wustl.edu  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.  
MGI:447533

Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 453.

Location/Qualifiers

## FEATURES

source

```
1. .459
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:722037"
/clone_lib="Soares mouse lymph node NbMLN"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Site.1: Not I - oligo(dt) primer
1st strand cDNA was primed with a Not I - oligo(dt) primer
5',
3'; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M. Fatima Bonaldo."
```

## BASE COUNT

107 a 126 c 110 g 116 t

## ORIGIN

Query Match 67.7%; Score 17.6; DB 30; Length 459;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GCCTCGAGTGTGAAAGCAAGATG 26

Db 197 GGCTGGAGTGTCAAAATCAAGATG 220

```
RESULT 6
AA271856 310 bp mRNA EST 26-MAR-1997
LOCUS v073h12.r1 Soares mouse 3NME12.5 Mus musculus cDNA clone
DEFINITION IMAGE:762695 5' similar to SW:RS27_HUMAN P42677 408 RIBOSOMAL
PROTEIN S27 ;, mRNA sequence.
ACCESSION AA271856
NID 91910185
VERSION AA271856.1 GI:1910185
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 310)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steproe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
TITLE Unpublished (1996)
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1402102.
COMMENT
```

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mousestewatson.wustl.edu  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.

MGI:463615  
putative full length read  
vector to vector length is 320  
Seq primer: -28ml3 rev2 ET from Amersham.

## FEATURES

source

```
1. .310
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:762695"
/clone_lib="Soares mouse 3NME12.5"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Site.1: Not I; Site.2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dt) primer
5',
3'; on total mouse RNA (provided by Minoru Ko, Wayne
State Univ.); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

## BASE COUNT

89 a 80 c 73 g 68 t

## ORIGIN

Query Match 67.7%; Score 17.6; DB 30; Length 310;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GCCTCGAGTGTGAAAGCAAGATG 26

Db 216 GCCTACAGTGTGAAAGCAAGATG 239

```
RESULT 7
AA981298 432 bp mRNA EST 27-MAY-1998
LOCUS vx60h11.r1 Stragatene mouse macrophage (#937306) Mus musculus cDNA
DEFINITION clone IMAGE:1279653 5' similar to TR:014730 O14730 HOMOLOG OF THE
ASPERGILLUS NIDULANS SUDG GENE PRODUCT. ;, mRNA sequence.
ACCESSION AA981298
NID 93159834
VERSION AA981298.1 GI:3159834
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 432)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steproe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
TITLE Unpublished (1996)
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2285238.
COMMENT
```

Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mousestewatson.wustl.edu  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (infoimage.llnl.gov) for further information.  
MGI:671453  
Possible reversed clone: similarity on wrong strand



Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 402.

## FEATURES

source

Location/Qualifiers  
1..432  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="19p12-p13.1"  
/clone="IMAGE:127953"  
/tissue="Stratagene mouse macrophage (#937306)"  
/dev\_stage="WBH1-3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: blood; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
0.19g dt. WBH1-3 cell line. Average insert size: 1.5 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GATTTCGCGACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 148 a 76 c 100 g 108 t  
ORIGIN

Query Match 67.7%; Score 17.6; DB 40; Length 432;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
0Y 3 GCCTGAGGTCGAAAGCAAGATG 26  
|||||  
Db 296 GCCTGAGGATGAAAGAGAGATG 319

## RESULT 8

LOCUS

DEFINITION

AI510391 409 bp mRNA EST 12-MAR-1999  
mp99c11.y1 Soares 2MBMT Mus musculus cDNA IMAGE:577364 5'  
similar to TR:014730 014730 HOMOLOG OF THE ASPERGILLUS NIDULANS  
SUDG GENE PRODUCT. ; mRNA sequence.

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 409)  
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R., and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Jun 5, 1998 this sequence version replaced gi:3189086.

Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone  
correct orientation)  
Seq primer: 40RP from Gibco.

## FEATURES

source

1..409  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/map="957B12"  
/clone="IMAGE:577364"

/clone\_lib="Soares 2MBMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pTR73-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - 0.19g(dt) primer [5,  
TGTTCACCAATCTGAAGTCGAGCGCGCGTTTCTTTTCTTTTCTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pTR73 vector. RNA  
provided by Dr. Bertrand Jordan. Library went through two  
rounds of normalization, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 145 a 73 c 82 g 109 t  
ORIGIN

Query Match 67.7%; Score 17.6; DB 47; Length 409;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
0Y 3 GCCTGAGGTCGAAAGCAAGATG 26  
|||||  
Db 325 GCCTGAGGATGAAAGAGAGATG 348

## RESULT 9

LOCUS

DEFINITION

AI608046 493 bp mRNA EST 21-APR-1999  
va05d11.y1 Soares mouse lymph node NbMLN Mus musculus cDNA clone  
IMAGE:722037 5', mRNA sequence.

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 493)  
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
Waterston, R., and Wilson, R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Jun 22, 1998 this sequence version replaced gi:3246994.

Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone  
correct orientation)  
Putative full length read  
vector to vector length is 523  
Seq primer: -40RP from Gibco  
High quality sequence stop: 466.

## FEATURES

source

1..493  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:722037"

```

/clone_lib="Soares mouse lymph node NbMLN"
/sex="male"
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT713D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15',
```

Query Match Best Local Similarity 83.3%; Pred. NO. 1.7e+02; Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 GCCTCAGGTCGAAGAACAAGATG 26  
 ||| | | | | | | | | | | | | |  
 Db 198 GGCTGGAGGTCCAAATCAAGATG 221

RESULT 10  
 AI661202/c 347 bp mRNA EST 10-MAY-1999  
 LOCUS VA05d11.x1 Soares mouse lymph node NbMLN Mus musculus CDNA clone  
 DEFINITION IMAGE:722037 3 , mRNA sequence.  
 ACCESSION AI661202  
 MID g4764785  
 VERSION AI661202.1 GI:4764785  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 347)  
 Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T.,  
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
 Rutter,E., Kohn,S., Shin,T., Jackson,F., Cardenas,M., McCann,R.,  
 Waterston,R. and Wilson,R.  
 The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 On Mar 20, 1998 this sequence version replaced gi:280279.  
 Other ESTs: va05d11.y1  
 Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel.: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL, contact the  
 IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
 MG1:447353  
 This clone was previously sequenced on the 5' end only, this new  
 data is from the 3' end  
 High quality sequence stop: 328.  
 Location/Qualifiers  
 1..347  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /map="19p12.00; 19q13.2"  
 /clone="IMAGE:722037"  
 /clone\_lib="Soares mouse lymph node NbMLN"  
 /sex="male"  
 /tissue\_type="lymph node"

```

/dev stage="" 4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pUT3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'
TGTTACCAATCTGAAGTCGAGCGCCGCAGACTGTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      104 a          75 c          88 g          80 t
ORIGIN
Query Match           67.7%; Score 17.6; DB 49; Length 347;
Best Local Similarity 83.3%; Pred.No.1.ee+02;
Matches    20; Conservative   0; Mismatches     4; Indels     0; Gaps    0;

Oy       3 GCCTCGAGGTCGAAAGCAAAGTG 26
         | | | | | | | | | | | | | | | | | | | |
Db        329 GGCTGGAGGTCCAAAATCAAGATG 306

RESULT  11
LOCUS    NA3298              361 bp             mRNA                  EST                12-DEC-1996
DEFINITION SW3ICA768SK Brugia malayi infective larva cDNA (SNA94NL-Bml3)
ACCSSION  NA3298               gi:1167680
NID       NA3298.1            GI:1167680
KEYWORDS  EST.
SOURCE    Brugia malayi.
ORGANISM  Eukaryota; Metazoa; Nematoda; Secernentea; Spirurina; Spiruriida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 361)
AUTHORS   Blaxter,M.L., Raghavan,N., Ghosh,I., Guillian,D., Lu,W., Williams,S.A., Slatko,B. and Scott,A.L.
TITLE     Genes expressed in Brugia malayi infective third stage larvae Mol. Biochem. Parasitol. 77 (1), 77-94 (1996)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:692807.

Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK
High quality sequence stop: 433.
Location/Qualifiers
1..361
/organism="Brugia malayi"
/strain="TRS Labs"
/db_xref="taxon:6279"
/cclone="SW3ICA768"
/cloned_lib="Brugia malayi infective larva cDNA (SNA94NL-Bml3)"
/lab_host="XLI-Blue MRF"
/note=vector: lambda unizap XR; site_1: EcoR I; site_2: Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from third stage infective larvae of
Brugia malayi isolated from mosquitoes 10 days after
infection and converted to double stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNAPol I. The library had 1.6 x 10E6 independent
recombinants and average insert size was 900 base pairs.
The library was constructed by Wenhong Lu. The library is

```

Query Match	66.2%;	Score 17.2;	DB 41;	Length 612
Best Local Similarity	86.4%;	Pred. No. 2.7e+02;		

Matches 19: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

QY 5 CTCGAGTCCGAAAGCAGATG 26

Db 574 CCCGAGGTGGGAAAGCAAG 595

## RESULT 14

R85938/c

LOCUS R85938 423 bp mRNA EST 14-AUG-1995  
DEFINITION Y037906.s1 Soares adult brain N2B4HB55Y Homo sapiens cDNA clone  
IMAGE:180154 3', mRNA sequence.

## ACCESSION

R85938

NID 9944344  
VERSION R85938.1 GI:944344  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

## REFERENCE

1 (bases 1 to 423)  
Hillier, L., Clark, N., Dubuque, T., Eliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

## AUTHORS

The WashU-Merck EST Project  
Unpublished (1995)  
On Nov 4, 1993 this sequence version replaced.

TITLE  
JOURNAL  
COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1422

High quality sequence stops: 320

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 1422 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 320.

## FEATURES

source

1. .423

/organism="Homo sapiens"

/db\_xref="GDB:382679"

/db\_xref="taxon:9606"

/clone="IMAGE:180154"

/clone\_lib="Soares adult brain N2B4HB55Y"

/sex="Male"

/dev\_stage="55-year old"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: brain; Vector: p773D (Pharmacia) with a

modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAACTGGAGCGCCGCTTTTCTTTTCTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p773 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 53. Library constructed by Bento

Soares and M.Fatima Bonaldo. The adult brain RNA was

provided by Dr. Donald H. Gilden. Tissue was acquired

17-18 hours after death which occurred in consequence of a

ruptured aortic aneurysm. RNA was prepared from a pool of

tissues representing the following areas of the brain:

frontal, parietal, temporal and occipital cortex from the

left and right hemispheres, subcortical white matter,

basal ganglia, thalamus, cerebellum, midbrain, pons and

medulla."

## BASE COUNT

93 a 118 c 121 g 88 t 3 others

## ORIGIN

## Query Match

65.4%; Score 17; DB 23; Length 423;

Best Local Similarity 80.0%; Pred. No. 3.1e+02;

Matches 20: Conservative 0: Mismatches 5: Indels 0: Gaps 0:

QY 2 TGCCTGAGTCCGAAAGCAGATG 26

Db 419 TTCACCGAGCGCCGAAAGCAGATG 395

## RESULT 15

H80272/c

LOCUS H80272 443 bp mRNA EST 09-NOV-1995  
DEFINITION Y059601.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:230401 3', mRNA sequence.

## ACCESSION

H80272

NID 91058361  
VERSION H80272.1 GI:1058361  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

## REFERENCE

1 (bases 1 to 443)  
Hillier, L., Clark, N., Dubuque, T., Eliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

## AUTHORS

The WashU-Merck EST Project  
Unpublished (1995)  
On Apr 14, 1993 this sequence version replaced gi:837779.

TITLE  
JOURNAL  
COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 747

High quality sequence stops: 329

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 747 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 329.

## FEATURES

source

1. .443

/organism="Homo sapiens"

/db\_xref="GDB:3781497"

/db\_xref="taxon:9606"

/clone="IMAGE:230401"

/clone\_lib="Soares fetal liver spleen INFLS"

/sex="Male"

/dev\_stage="20 week-post conception fetus"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)

with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer

[5' AACGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],

double-stranded cDNA was ligated to Eco RI adapters

(Pharmacia), digested with Pac I and cloned into the Pac I

and Eco RI sites of the modified p773 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

## BASE COUNT

122 a 94 c 97 g 128 t 2 others

## Query Match

65.4%; Score 17; DB 24; Length 443;

Thu Sep 23 10:37:34 1999

us-08-671-757-1.rst

Page 9

Best Local Similarity 80.0%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 TGGCTGAGGTGGAAGCAGATG 26  
||| | ||||| ||||| |||||  
DB 206 TGGGTGAGGTCCAAAGGAGCTTG 182

Search completed: September 22, 1999, 16:59:01  
Job time: 321 sec

